

# 基于多线索的目标跟踪

张海青 李厚强

(中国科学技术大学电子工程与信息科学系,合肥 230027)

**摘要** 通过对人眼跟踪机制的研究,提出了一种新的基于多线索的目标跟踪方法。该方法采用串行结构处理多个视觉线索,首先按近邻原则产生若干候选目标,然后使用不同线索按优先级顺序逐次对候选目标进行筛选,得到的唯一候选目标再经过校正以获得更为准确的跟踪结果。该方法最大的特点是跟踪系统对环境和场景的变化有很强的自组织和自适应能力,系统内多个线索在跟踪过程中的竞争与协同使得跟踪具有强大的适应力和生命力。实验结果表明,该方法显著地提高了跟踪的鲁棒性和准确性。

**关键词** 目标跟踪 多视觉线索 串行结构 Mean Shift 算法

中图分类号:TP391.41 文献标识码:A 文章编号:1006-8961(2008)11-2187-10

## Object Tracking Through Multiple Cues

ZHANG Hai-qing, LI Hou-qiang

(Department of Electronic Engineering and Information Science, University of Science & Technology of China, Hefei 230027)

**Abstract** In this paper a novel approach for object tracking using multiple cues is presented, based on the investigation of human eye tracking. In this approach, multiple cues with serial structure are used in tracking. First, some candidate objects are found through near-neighbor principle. Then multiple cues with different priorities select the candidates in-order. At last, the only survive candidate is fined by the modified module to get more accurate tracking result. The most salient characteristic of this approach is the principles of self-organization and self-adaptation of the changing environment during tracking. Multiple cues compete and cooperate in the system, which make the tracking has strong adaptation and vitality. Experiments show the robustness and accuracy of the tracking algorithm.

**Keywords** object tracking, multiple cues, serial structure, Mean Shift algorithm

## 1 引言

目标跟踪是近年来计算机视觉领域中备受关注的方向,在视觉处理系统中,它是位于运动检测和目标分类之后、行为理解与描述之前的中级处理部分,在智能监控系统、虚拟现实、高级用户接口、运动分析和高效视频编码等方面有着广泛的应用前景。

简单而言,跟踪就是在图像序列中创建基于位置、色彩、纹理、形状、运动等有关特征的对应匹配问

题。随着卡尔曼滤波、Mean Shift 算法和粒子滤波等工具的应用,近年来跟踪算法得到了长足的发展。如基于 Mean Shift 的跟踪方法<sup>[1,2]</sup>通过构造目标的核密度表示模型,将目标跟踪转化为模式搜索问题,然后利用 Mean Shift 算法(一种基于核密度估计的无参快速模式匹配算法)得到最佳匹配结果,提高了跟踪的实时性;基于粒子滤波的跟踪<sup>[3]</sup>是一种基于贝叶斯统计计算的方法,该方法首先把跟踪问题转化为非线性非高斯的贝叶斯滤波问题,然后利用粒子滤波求解后验概率从而获得跟踪目标,特别适用于多目标和存在多假设目标时的跟踪。

**基金项目:**国家自然科学基金委员会与微软亚洲研究院联合资助项目(60672161)

**收稿日期:**2007-03-09; **改回日期:**2007-04-28

**第一作者简介:**张海青(1977~),男,工程师。中国科学技术大学电子工程与信息科学系硕士研究生。主要研究方向为计算机视觉、智能监控。E-mail: wjzhq@mail.ustc.edu.cn

目前仍然困扰跟踪的问题是如何进一步提高跟踪的鲁棒性和准确性,因为在许多应用领域中,需要在复杂环境和场景下完成跟踪任务,具体来说就是要求跟踪算法能够解决目标灵活运动、尺度变化、形状变化、环境光线的渐变和突变、目标遮挡和阴影、目标类似干扰等问题。

然而,目前大部分的跟踪方法都是基于单一视觉线索的跟踪方法,如文献[1]、[2]只利用了目标的颜色特征,文献[4]只利用了纹理特征,文献[5]只利用了形状和轮廓特征等。单一视觉线索跟踪方法对环境变化敏感,鲁棒性不高,因为当环境和场景变化时,单一特征或线索容易丢失,使跟踪失败或变得非常困难。为了提高跟踪算法的鲁棒性,根本的方法是同时利用多种线索对目标进行跟踪。

Democratic Integration<sup>[6]</sup>是一种多线索的目标跟踪方法,该方法采用并行结构处理多个线索,首先对每个线索计算相应的显著性图,然后根据各个线索对跟踪结果重要性的不同计算融合后的显著性图,最后通过检测极大值的方法获取跟踪目标并对线索进行更新。该方法结构简单、易于实现,缺点是由于采用并行结构处理多个线索,使得跟踪过程需要对每个线索处理一次,即使是那些无用(在某种环境和场景下失效)的线索,增加了系统的负担,不符合人眼跟踪的自组织性;另外,由于对每个线索处理所使用的方法过于简单,使得最后跟踪的结果不够准确,只能作为视觉搜索的预注意(神经生理学概念)结果。多线索粒子滤波目标跟踪方法<sup>[7]</sup>是在统计意义上的多线索跟踪方法,该方法通过构造多线索的联合观测概率密度来估计目标的后验密度,从而得到跟踪结果,其使用多线索的方法与 Democratic Integration 方法相同,因此有同样的缺点。

本文通过对人眼跟踪机制的研究,提出了一种新的基于多线索的目标跟踪方法。该方法采用串行结构处理多个线索,首先按近邻原则产生若干候选目标,然后使用不同线索按优先级顺序逐次对候选目标进行筛选,得到的唯一候选目标再经过校正以获得更为准确的跟踪结果。该方法最大的特点是跟踪系统对环境和场景变化有很强的自组织和自适应能力,系统内多个线索在跟踪过程中的竞争与协同使得跟踪具有强大的适应力和生命力。实验结果表明,该方法显著地提高了跟踪的鲁棒性和准确性。

## 2 串行结构多线索目标跟踪思想

多线索目标跟踪的难点在于,首先如何理解人

眼目标跟踪的机制及其使用多线索的方法,其次如何根据这一机制建立一个合理的跟踪框架,实现一个鲁棒而准确的跟踪算法。

### 2.1 人眼目标跟踪机制的理解

由于人类的视觉系统极为复杂,目前人们只对视觉信息处理的整体流程有一定的了解,很多的视觉现象和特性还没有被清楚地解释,尽管如此,一些研究知识已经成功地运用到图像的分析处理过程中,并对指导计算机视觉研究起了很大的影响。本节试图从这些已知的研究成果中获得对人眼目标跟踪的理解。

视觉注意力<sup>[8]</sup>是人类视觉系统的特性之一,注意力是一个神经生理学的概念,表示人在仔细观察一个物体时精神集中的能力。视觉生理学方面的研究表明,当图像呈现在人面前时,只有少部分的信息能够进入短时记忆并得到进一步的处理,注意力在这个过程中起了类似于过滤器的作用。注意力的选择性使得人们可以只把图像中吸引注意力的部分呈现给用户,而对用户的视觉体验没有很大的影响。这一机制的内在原因是感光细胞在视网膜上分布的极为不均匀性和视觉加工系统是资源有限的容量系统,人类对信息的选择依赖于中枢对视觉器官的反馈和调节。

认知心理学中经典的视觉搜索任务<sup>[9]</sup>研究了人脑的视觉搜索策略,在联合特征任务中,对于人脑是如何将不同特征联合在一起的这一问题一直存在着争论,Friedman-Hill 和 Wolfe<sup>[10]</sup>提出了一种视觉搜索范式,称为子集搜索。在子集搜索中首先根据某一特征分组形成一个子集,然后在同一特征组中找到不同的项目,即靶刺激,这个过程没有涉及特征捆绑,只有按相同特征进行分组的过程,这一理论得到了许多实验和理论<sup>[11]</sup>的支持。

自组织现象在自然界和人类社会中普遍存在,人眼目标跟踪系统能够根据跟踪中出现的不同情况,按照相互默契的某种规则,各尽其责而又协调、自动地形成内在有序结构,这就是人眼的自组织性,这种自组织性使得人眼跟踪具有很强的适应环境的能力和旺盛的生命力。

另外认知心理学的研究<sup>[12]</sup>说明了人眼视觉加工一般可以分为两个阶段,视觉机制在早期阶段对整个视域执行加工,以获得感兴趣区域,在后期阶段只对有限视域中的对象进行精细加工,以完成视觉任务,从某种意义上说,这也反映了人类视觉的注意力特性。

以上的认识为理解人眼目标跟踪机制提供了有益的启示:(1)人眼跟踪系统是一个复杂的自组织系统,具有很强的适应环境的能力;(2)在跟踪中利用人眼选择性的注意力分配特性,可以分解信息处理过程;(3)子集搜索策略为解决跟踪中使用多个线索的问题提供了一个合理而有效的方法;(4)可以把目标跟踪过程分为感兴趣区域提取和精细校正两部分,以减轻信息处理过程的负担,提高跟踪的准确性。

### 2.2 基于串行结构的多线索目标跟踪框架

根据上一节对人眼目标跟踪机制的理解,提出一个多线索目标跟踪框架,目的是使算法在复杂环境和场景变化时也能对目标进行鲁棒的跟踪。框架把跟踪过程分解为候选目标的提取和校正两部分,在候选目标提取过程中,各线索采用自组织的串行结构逐步过滤候选目标,直至得到唯一候选,校正过程对候选目标进一步加工,使得最终的跟踪结果更为准确。系统最显著的特点是它能够通过反馈机制完成自组织的线索联合和自适应的线索更新,保证系统结构对环境和场景变化的最优性。

基于串行结构的多线索目标跟踪框架如图 1 所示。视频图像首先进入抽样过程,由近邻原则(即相邻两帧目标位移和尺度变化不会太大)在目标预测位置附近得到多个抽样候选目标;之后进入多线索过滤阶段,此过程使用目标的运动、色彩、纹理、形状、预测等信息逐步对候选目标进行过滤,一旦获得唯一结果则终止线索过滤,进入目标校正过程;目标校正是在唯一候选目标位置附近搜索局部极值的过程,用来准确定位目标;校正后的结果作为跟踪输出,同时将其反馈到线索更新单元,以对各线索参考值和优先级分别进行调整,用以指导后续的跟踪,线索更新单元维持一张线索表(图 1 左下所示),分别记录各线索参考值和对应的优先级,线索表反映了当前的环境和场景知识,其中优先级高的线索在多线索过滤阶段处于较前的位置,预测信息的优先级始终最低以保证多线索过滤有唯一的输出;初始化阶段时,首先通过手工选择或根据目标检测得到跟踪目标,然后计算跟踪目标各线索的参考值并赋予一定的优先级,以产生一个初始的线索表,指导下一步的跟踪。

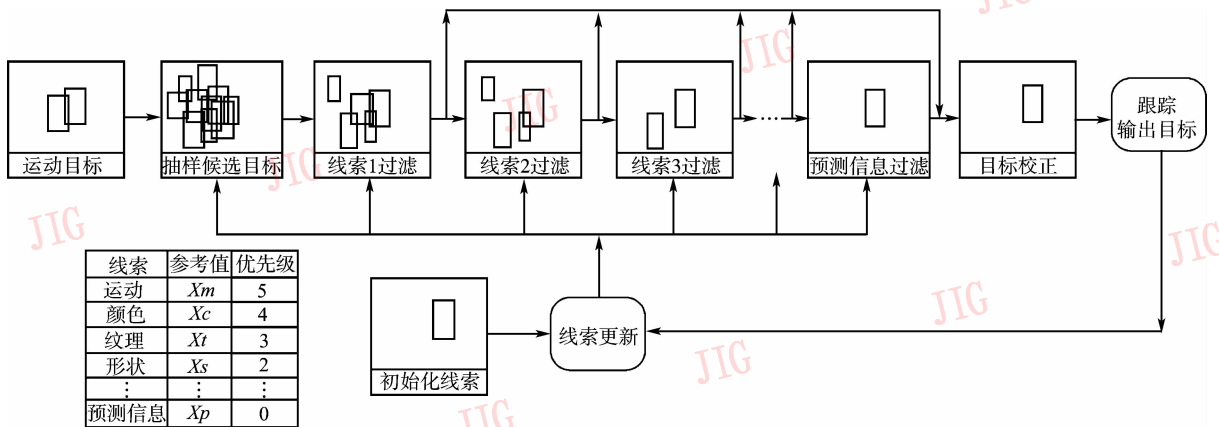


图 1 基于串行结构的多线索目标跟踪框架  
Fig. 1 General framework for object tracking with serial multiple cues

## 3 串行多线索目标跟踪方法

### 3.1 目标的多视觉线索

目标线索的提取与表达是多线索目标跟踪的基础,这里的线索指的是目标广义上的所有特征或信息,即区别于其他目标的所有可利用信息。对于视频图像而言,除了常见的图像特征,如色彩、纹理、形状、运动等外,其他的信息如目标的局部特征等也可作为跟踪的线索。

颜色特征之所以选择在目标跟踪中应用广泛的

视觉线索,其主要原因在于颜色往往和图像中所包含的物体或场景十分相关,且提取与表达较为简单,适用于实时跟踪的要求,此外与其他特征相比,颜色特征对图像本身的尺寸、方向、视角的依赖性较小,对于目标尺度变化和形变时的跟踪有较高的鲁棒性,但颜色对光线变化敏感,因此在光线突变时颜色特征容易丢失。

纹理特征是一种不依赖于颜色或亮度的反映图像中同质现象的视觉信息,它是所有物体表面共有的内在特性。纹理特征包含了物体表面

结构组织排列的重要信息以及它们与周围环境的联系,因此,纹理特征的使用可以克服环境光线变化对跟踪带来的影响,但目标被遮挡时纹理特征也将消失。

形状特征是目标跟踪中的另一重要线索,但不同于颜色或纹理等底层特征,形状特征的表达必须以对图像中物体或区域的分割为基础。由于当前的技术无法做到准确而鲁棒的自动分割,这给利用形状特征进行跟踪带来了难度,但是针对不同的应用,选用简单的形状表达仍可提高跟踪的鲁棒性。

运动特征是视频序列中特有的目标信息,在复杂场景下(如出现目标遮挡、类似干扰等),仅仅利用颜色或纹理特征进行跟踪是不够的,此时目标的运动特征将是跟踪的重要线索,利用它可以获得正确的跟踪目标。

一个优秀的跟踪系统必然需要使用目标的多个视觉线索,只有这样才能得到更加准确的跟踪结果。

### 3.2 候选目标的提取过程

候选目标的提取是运用子集搜索策略通过不同线索对感兴趣目标逐渐过滤和筛选直至最后得到唯一结果的过程。这个过程首先产生一系列抽样候选目标;接着依据注意理论淘汰一些不受关注的抽样,此步骤实际利用了目标的运动信息;然后根据线索表依次选择不同的线索对候选目标进行过滤,获得唯一候选目标后终止后续线索过滤,结果送入目标校正过程。

#### 3.2.1 抽样候选的产生

抽样候选目标产生的依据是近邻原则,即相邻两帧目标位移和尺度变化不会太大,因此可以假设当前帧目标位于目标预测位置附近,尺度与预测大小接近。如果用矩形框来表示跟踪目标,如图 2(a)所示,那么假设当前帧中表示候选目标的矩形框的中心服从 2 维高斯分布,尺度服从 1 维高斯分布是合理的,高斯分布的均值分别为预测目标矩形的中心和尺度,方差大小反映了目标位移和尺度变化的速率。当前帧目标位置和尺度的预测使用卡尔曼滤波器。

矩形中心坐标的 2 维高斯分布可由 Box-Muller 方法<sup>[13]</sup>产生,令随机变量  $x, y$  是服从  $[0, 1]$  的均匀分布,那么式(1)表示的随机变量  $u, v$  即服从均值为  $\mu$ , 方差为  $\sigma$  的 2 维高斯分布:

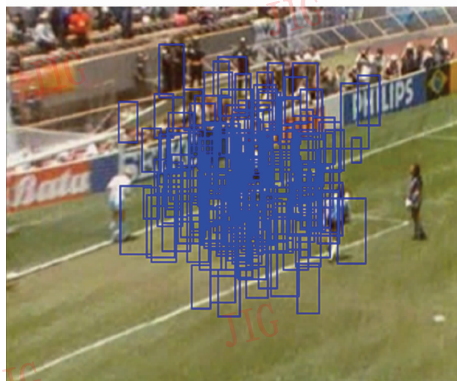
$$u = (\sqrt{-2\ln x} \cos(2\pi y)) \times \sigma + \mu \quad (1)$$

$$v = (\sqrt{-2\ln x} \sin(2\pi y)) \times \sigma + \mu$$

对矩形框的位置和尺度进行抽样,产生当前帧的抽样候选目标,用矩形框表示,如图 2(b)所示。只要抽样数足够多,可以保证在这些抽样中总有一个与真实目标接近的矩形框,后续步骤就是在这些抽样中逐步挑选出与真实目标最接近的候选目标矩形框,然后进行校正得到最终的跟踪结果。



(a) 前一帧的跟踪目标



(b) 当前帧 200 个抽样矩形

图 2 当前帧的候选目标

Fig. 2 Candidate objects of current frame

#### 3.2.2 注意区域的选择

按注意理论,在视觉搜索中人眼首先会根据任务的性质注意到一些感兴趣的区域,然后进行处理。在目标跟踪中,相邻帧存在变化的地方首先是感兴趣的区域,真实目标由于运动将包含在这些区域中,因此由此可以淘汰一些候选矩形框,这些矩形框对应的区域相邻帧变化不大。这一过程实际上利用了跟踪目标存在运动这一信息,因此这一过程也可理解为是利用目标运动信息对候选目标进行过滤的过程。

具体方法为,首先由式(2)产生  $t$  时刻两帧差的

二值化图  $M(x, t)$ , 式中  $v(x, t)$  为  $t$  时刻  $x$  处的像素值,  $\tau_m$  为二值化阈值:

$$M(x, t) = H(|v(x, t) - v(x, t - 1)| - \tau_m) \quad (2)$$

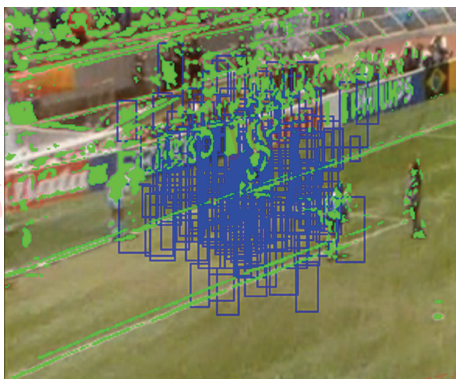
$$H(z) = \begin{cases} 0 & z \leq 0 \\ 1 & z > 0 \end{cases}$$

然后按(3)式计算每一个抽样候选的累计帧差值,  $V(i, t)$  为  $t$  时刻第  $i$  个抽样的累计帧差值,  $num$  为第  $i$  个抽样对应矩形框的像素数:

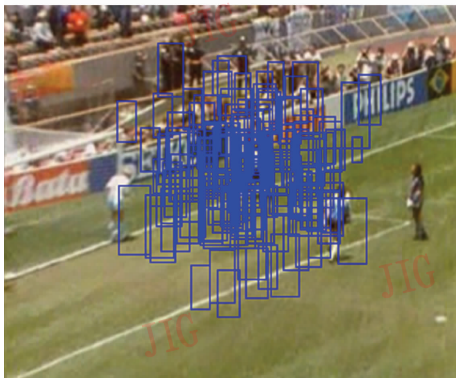
$$V(i, t) = \sum_{j=1}^{num} M(x_j, t) \quad (3)$$

最后累计帧差值大于某一阈值的候选目标保留, 其他的则删除。另外, 此过程实际利用了目标存在运动这一假设, 为了不排除存在目标静止的情况, 在此模块的输出中, 增加一个表示前一帧目标跟踪结果的矩形框, 它与保留的候选目标矩形框一起进入下一步的过滤阶段。

由式(2)可以得到两帧差的二值化图, 图 3(a) 中绿色表示运动区域, 注意力首先集中于此; 图 3(b) 为经过注意区域选择后保留的候选目标。



(a) 相邻帧的变化



(b) 保留的候选目标

图 3 注意区域的选择

Fig. 3 Remaining objects after visual attention

### 3.2.3 颜色特征过滤

使用颜色直方图作为目标的颜色特征, 计算颜色直方图首先进行颜色量化, 将颜色空间划分成若干小的颜色区间, 每个小区间成为直方图的一个 bin; 然后通过计算颜色落在每个小区间内的像素数量得到颜色直方图。为了使算法简单实用, 使用 RGB 颜色空间, 并将其均匀划分成  $8 \times 8 \times 8$  个 bins, 参考目标区域的颜色特征可由下式计算:

$$p_u(\mathbf{y}) = c \cdot \sum_{i=1}^n \delta[b(\mathbf{x}_i) - u] \quad (4)$$

式中,  $c$  为归一化系数, 函数  $b: R^2 \rightarrow \{1, \dots, m\}$  计算每个像素  $\mathbf{x}_i$  的颜色值,  $u = 1, \dots, m$  表示颜色直方图的  $m$  个颜色小区间, 用同样的方法可以求得候选目标区域的颜色特征  $q_u(\mathbf{y})$ 。

用 Bhattacharyya 系数<sup>[14]</sup>来描述参考目标和候选目标的相似程度:

$$\rho[\mathbf{p}, \mathbf{q}] = \sum_{u=1}^m \sqrt{p_u q_u} \quad (5)$$

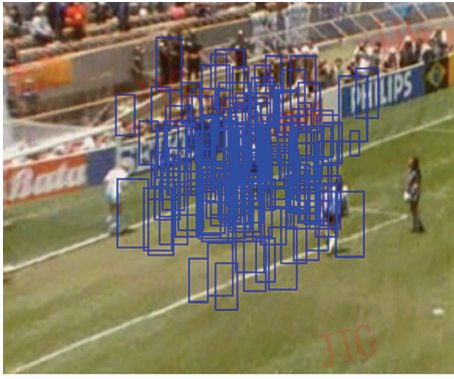
使用颜色特征对候选目标进行过滤的方法是, 当候选目标与参考目标的颜色相似度大于某一阈值时, 候选目标保留, 其他的则删除。如果保留目标只剩下一个, 则跳过后面的特征过滤直接进入目标校正过程; 如果候选目标全部被删除, 说明使用颜色特征失效, 此时使模块的输出等于输入; 其他情况时保留的目标进入下一特征过滤, 直至剩下唯一的候选目标。

图 4(a) 为使用颜色特征过滤前保留的候选目标, 图 4(b) 为经过颜色特征过滤后输出的候选目标。从图中可以看到, 只有少数的具有相似颜色特征的候选目标得以保留, 这样就大大减少了下一步处理的负担。

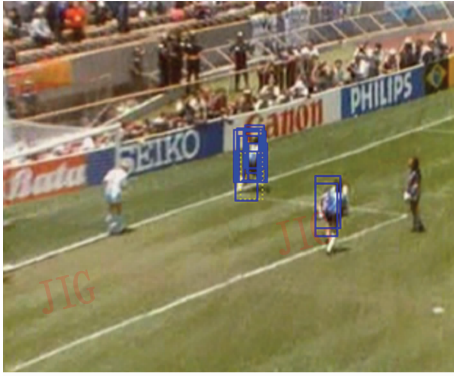
### 3.2.4 纹理特征过滤

目标的纹理特征使用纹理谱<sup>[15]</sup>的方法获得, 图像中每个像素(除边界像素外)都有 8 个相邻像素, 像素的局部纹理特征可以根据其  $3 \times 3$  的邻域内的像素提取, 这个局部纹理特征称为最小的纹理单元。定义纹理单元为包含 8 个像素的一个集合:  $\mathbf{TU} = \{E_1, E_2, \dots, E_8\}$ , 其中,  $E_i$  的值由下式确定,  $v_0$  为被处理像素,  $v_i$  为其邻域中的像素,  $\Delta$  代表一个很小的正值:

$$E_i = \begin{cases} 0 & v_i \leq v_0 - \Delta \\ 1 & v_0 - \Delta < v_i \leq v_0 + \Delta \\ 2 & v_i > v_0 + \Delta \end{cases} \quad (6)$$



(a) 输入的候选目标



(b) 输出的候选目标

图 4 颜色特征过滤

Fig. 4 Remaining objects after color filtering

整幅图像中不同纹理单元分布情况的统计反映了图像整体的纹理信息。定义所有纹理单元出现的频率为纹理谱,从而可生成纹理谱像,其中横坐标为纹理单元标记号,纵坐标为出现的频率。用下式计算参考目标和候选目标的相似程度:

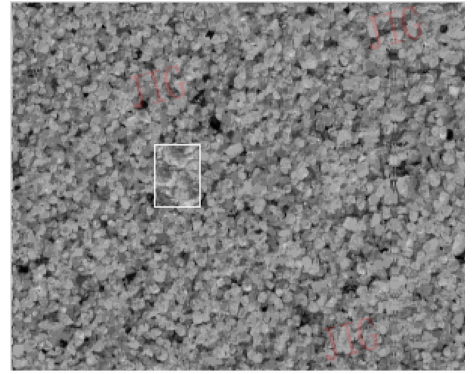
$$\rho[\mathbf{T}p, \mathbf{T}q] = \sum_{u=1}^n \sqrt{\mathbf{T}p_u \mathbf{T}q_u} \quad (7)$$

式中,  $\mathbf{T}p$  和  $\mathbf{T}q$  分别为参考目标和候选目标的纹理谱,  $u$  为纹理单元标记,  $n$  为纹理单元数。

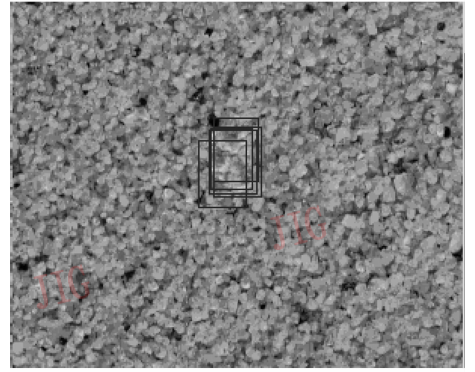
使用纹理特征对候选目标进行过滤的方法与使用颜色特征时相似,纹理相似度大于某一阈值的候选目标保留,其他的则删除。如果保留目标只剩下一个,则跳过后面的特征过滤进入目标校正过程;如果候选目标全部被删除,说明跟踪使用纹理特征失效,此时使模块的输出等于输入;其他情况时保留的目标进入下一特征过滤。

图 5(a) 为人工合成图像,其中用矩形框表示的区域的纹理与整幅图像的纹理不同,但颜色相似,此时如果需要跟踪该区域只能使用纹理信息;图 5(b) 为经过纹理特征过滤后输出的候选目标。从图中可

以看到,使用本节的纹理特征过滤方法可以得到理想的候选结果,该纹理特征方法比其他方法(如共发矩阵法等)简单有效。



(a) 不同的纹理图像



(b) 相似的纹理区域

图 5 纹理特征过滤

Fig. 5 Remaining objects after texture filtering

### 3.2.5 运动预测信息过滤

该过程使用了目标的运动预测信息对候选目标进行过滤,当前面使用的特征未能得到唯一的候选目标时,最后使用目标的运动预测信息进行过滤,该过程产生唯一的候选目标,目标的运动预测信息用卡尔曼滤波器<sup>[16]</sup>获得。

用矩形框描述目标的位置和大小,目标的运动预测信息由运动矢量描述,因此滤波器状态变量用表征该目标矩形框的一对对角顶点的位置、速度和加速度表示:

$$\mathbf{x} = (x_1, y_1, x_2, y_2, v_{x_1}, v_{y_1}, v_{x_2}, v_{y_2}, a_{x_1}, a_{y_1}, a_{x_2}, a_{y_2})^T$$

由于对两个顶点的滤波过程相同,可以将其分解为两个滤波器。

根据目标运动的规律建立状态转移方程,假设目标在相邻两帧间的运动近似为匀加速直线运动,因此状态转移方程如下式所示:

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \boldsymbol{\omega}_k \quad (8)$$

式中,  $\mathbf{x}_k = (x_{1k}, y_{1k}, v_{x_{1k}}, v_{y_{1k}}, a_{x_{1k}}, a_{y_{1k}})^T$  为一个顶点的状态变量,  $\boldsymbol{\omega}_k$  为状态转移噪声,  $\mathbf{A}_k$  为二阶牛顿动力学方程决定的状态转移矩阵,即

$$\mathbf{A}_k = \begin{bmatrix} 1 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

因为只能观测到目标的位置,所以观测方程描述的是顶点的位置,观测方程如式(9)所示:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \quad (9)$$

式中,  $\mathbf{z}_k = (z_{x_{1k}}, z_{y_{1k}})^T$  为观测矢量,  $\mathbf{v}_k$  和  $\mathbf{H}_k$  分别为观测噪声和观测矩阵,即

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

根据卡尔曼滤波理论,可以获得两个顶点的预测值,由此确定跟踪目标的预测位置和大小。

使用运动预测信息对候选目标进行选择是比较候选矩形框与预测矩形框距离的过程,距离最近的那个矩形框保留,作为多线索过滤后唯一的输出,距离按下式计算:

$$d = w \cdot d_{\text{location}} + (1 - w) \cdot d_{\text{scale}} \quad (10)$$

$$d_{\text{location}} = \sqrt{(x_{o_1} - x_{o_2})^2 + (y_{o_1} - y_{o_2})^2}$$

$$d_{\text{scale}} = \left| \sqrt{s_1} - \sqrt{s_2} \right|$$

式中,  $d_{\text{location}}$  和  $d_{\text{scale}}$  分别为位置距离和尺度距离,  $w$  为权重,  $(x_{o_1}, y_{o_1}), (x_{o_2}, y_{o_2})$  为两矩形框中心坐标,  $s_1, s_2$  为矩形框的面积。

图 6 中蓝色矩形框为使用目标的运动信息过滤前保留的候选目标,黄色矩形框为利用跟踪目标的运动信息过滤后得到的唯一候选目标,它是经过多线索逐步挑选后得到的与真实目标最近的一个候选目标,将其作为粗略的跟踪结果,并送入校正步骤。

以上使用的视觉线索对于大多数跟踪已经足够,但对于特别复杂的情况,可以在使用运动预测信息前加入更多的线索,以达到鲁棒跟踪的要求。

### 3.3 候选目标的校正过程

目标校正是在多线索过滤后得到的唯一候选目标附近寻找局部极值的过程,目的是获得更为准确



图 6 运动预测信息过滤

Fig. 6 Remaining objects after motion filtering

的跟踪结果。Mean Shift 算法<sup>[2]</sup>是一种基于核密度估计的无参快速模式匹配算法,可以得到局部最佳的匹配结果,目标模型采用引入核密度函数的归一化特征分布,如下式所示,

$$p_u(\hat{\mathbf{y}}_0) = C_h \sum_{i=1}^n k \left( \left\| \frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h} \right\|^2 \right) \delta[b(\mathbf{x}_i) - u] \quad (11)$$

其中,  $p_u(\hat{\mathbf{y}}_0)$  为引入核密度函数后候选目标的颜色特征,  $C_h$  为归一化系数,  $k(\cdot)$  为 Epanechnikov 核函数对应的轮廓函数。特征使用此时优先级最高的特征,为了使 Mean Shift 矢量计算简单,采用 Epanechnikov 核函数,因为其对应的轮廓函数的导数为常数,另外为了能够得到尺度上的最佳匹配,选用可变的核函数宽度。

假设校正时线索表中优先级最高的线索为颜色特征,那么用 Mean Shift 算法对候选目标进行校正的过程可以简单归纳如下:

(1) 初始化当前帧中目标的位置为  $\hat{\mathbf{y}}_0$ ,按式(11)和式(5)分别计算颜色特征和相似度;

(2) 计算权值  $\{w_i\}$

$$w_i = \sum_{u=1}^m \delta[b(\mathbf{x}_i) - u] \sqrt{\frac{q_u}{p_u(\hat{\mathbf{y}}_0)}}$$

(3) 使用 Mean Shift 算法计算目标新位置  $\hat{\mathbf{y}}_1$

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^n \mathbf{x}_i w_i}{\sum_{i=1}^n w_i}$$

(4) 更新颜色特征和相似度,当更新后的相似度小于更新前的相似度时,  $\hat{\mathbf{y}}_1 \leftarrow (\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)/2$  并返回上一步,直到更新后的相似度大于更新前的相似度;

(5) 如果  $\|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\| < \varepsilon$  则停止,  $\varepsilon$  为很小的值,否则置  $\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1$  并返回步骤 1。

利用其他特征或线索进行目标校正时,采用类似的方法。

图 7 中黄色矩形框为利用颜色信息进行校正后得到的最后的跟踪结果,与图 6 中得到的候选目标相比,其位置和尺度都改变了,可以看到校正后得到的跟踪结果更加准确。



图 7 校正后得到的跟踪结果

Fig. 7 The tracking result after adjustment

### 3.4 线索的初始化及更新

初始化阶段时,首先通过手工选择或根据目标检测得到要跟踪的目标,然后计算目标各线索的参考值并赋予一定的优先级,优先级的范围为 0~5,每一个线索初始化一个可信度用于参考特征值的更新,可信度一般为 1~10。更新阶段时,包括对线索参考值和优先级的自适应调整。

线索优先级用于调整过滤过程中使用各线索的先后顺序,其更新的方法是当得到当前帧的跟踪目标后,考查跟踪过程中所使用的线索,对那些在过滤候选目标过程中起作用的线索的优先级增加 1(最大为 5),没有起作用的减去 1(最小为 0),起作用线索定义为能够过滤掉部分候选目标(即有输出并且输出候选数小于输入数)的线索。由于人眼首先注意运动存在变化的区域,所以定义 3.2.2 节中的目标运动信息的优先级始终最高,另外只有运动预测信息过滤才能保证唯一的输出,因此定义 3.2.5 节中的运动预测信息的优先级始终最低。

线索(或称广义特征)参考值是跟踪目标的真实表达,其更新使用混合模型法。每一个特征(如颜色、纹理等)的参考值用 3~5 个不同的分特征(如颜色 1、颜色 2、颜色 3 等)的值混合表

示,每一个分特征有一个可信度,可信度最高的分特征值为当前真实参考特征值,其他的为虚假特征值,虚假特征值的出现是由于发生了目标遮挡、光线突变等情况。跟踪得到的目标的某一特征(如颜色)分别与该特征对应的每一个分特征(如颜色 1、颜色 2、颜色 3 等)匹配,匹配使用相似度来衡量,相似度的计算采用式(5)和式(7),相似度超过阈值且为最大的那个分特征为匹配分特征。有匹配时,对匹配分特征的可信度增加 1,对应的值得到更新,其他的可信度减去 1,值保持不变;无匹配时,可信度最小的那个分特征去除并用跟踪得到的目标的特征代替。匹配时,用下式对分特征值进行更新, $i$  为对应的匹配分特征。

$$\mathbf{x}_k^i = \mathbf{x}_{k-1}^i + \alpha(\mathbf{x}_k - \mathbf{x}_{k-1}^i) \quad (12)$$

其中, $\mathbf{x}_{k-1}^i$  和  $\mathbf{x}_k^i$  分别为  $k-1$  和  $k$  时刻第  $i$  个分特征的值, $\mathbf{x}_k$  为跟踪得到的目标的特征值, $\alpha$  为更新率。使用混合模型法对特征参考值进行更新,可以有效地解决由遮挡、光线突变或误跟踪时产生的虚假目标特征对真实参考特征的影响,从而提高跟踪的鲁棒性。

## 4 实验结果和分析

为了说明并检验本文提出的基于串行结构的多线索目标跟踪方法有很强的鲁棒性和准确性,能够解决复杂环境和场景下的跟踪问题,使用以下视频序列进行目标跟踪实验。实验在 PC 机(Celeron CPU 1.70G/WindowsXP)上运行,视频图像帧的大小为  $300 \times 245$  pixels。初始化时,各线索按先后顺序分别为运动信息、颜色信息、纹理信息和运动预测信息,优先级都为 3,特征可信度为 5,抽样数为 100。

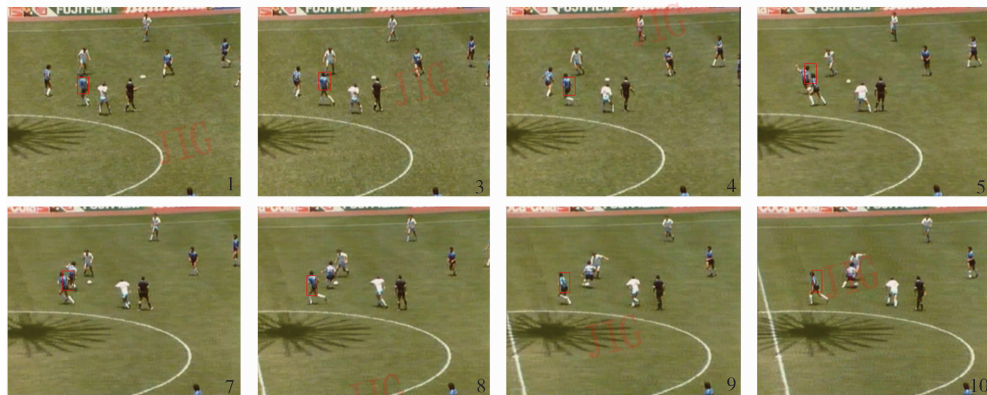
图 8(a)中足球运动员由于自身和摄像机的运动,产生了位置、形状和尺度的变化。采用本文的方法解决在此情况下的跟踪问题时,首先对目标进行位置和尺度抽样,以适应目标的位置和尺度变化,这种方法对目标运动不可预测的情况尤为适用;然后多特征过滤获得唯一候选目标,由于颜色特征对形状和尺度变化有较高的鲁棒性,因此在这种情况下,跟踪算法实际只用了目标的运动、颜色和预测信息;最后通过校正获得了很好的跟踪结果,如图 8 矩形框所示。



(a) 目标位置、形状和尺度变化时的跟踪



(b) 发生遮挡时目标的跟踪



(c) 目标存在类似干扰时的跟踪

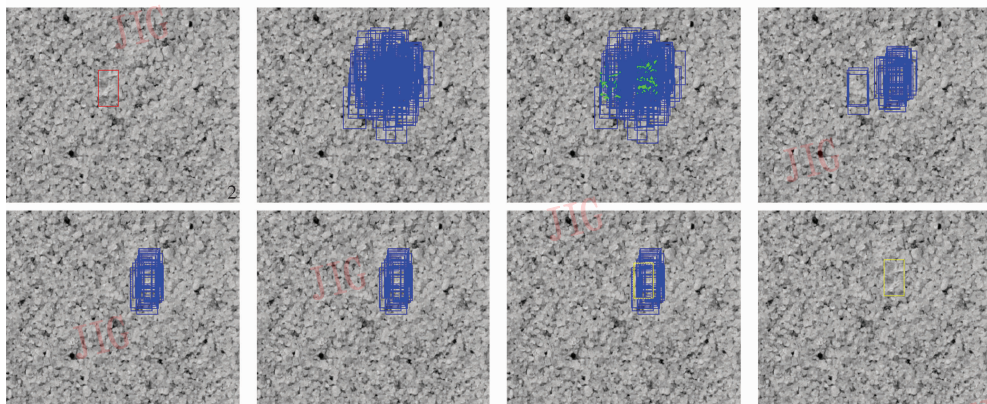


图 8 基于串行结构的多线索目标跟踪方法的跟踪结果

Fig. 8 Tracking results with serial multi-cue methods

图 8(b) 中第 7 帧和第 8 帧目标被遮挡, 此时目标的颜色和纹理特征都消失, 不能利用它们继续进行跟踪, 此时目标的运动预测信息为跟踪提供了有效的线索, 利用它得到了较为合理的跟踪结果。

图 8(c) 中跟踪运动员穿蓝色上衣, 附近有一个与他颜色、纹理、尺度都相近的运动员, 而且两运动员出现交叉, 容易产生类似干扰。此时仅仅利用颜色和纹理特征不足以对目标进行正确跟踪, 然而本文方法合理地利用了目标的其他线索, 即运动预测信息, 有效地解决了类似干扰问题, 体现了采用多线索进行跟踪的优势。

图 8(d) 跟踪红色矩形框区域, 由于跟踪区域的颜色与整幅图像的颜色相似只有纹理不同, 此时颜色特征失效, 将主要使用纹理信息对目标进行跟踪, 由实验结果看到, 算法可以得到正确的跟踪结论。图 8(d) 中由上到下, 从左至右依次为: 需要跟踪的目标、抽样候选目标、注意区域、注意区域选择后保留的目标、颜色特征过滤后保留的目标、纹理特征过滤后保留的候选目标、运动信息过滤后得到的唯一候选目标(黄色矩形框)、利用纹理信息进行校正后得到的最后的跟踪结果。

通过实验可以看到, 基于串行结构的多特征目标跟踪方法有效地利用和组织了目标的多个线索, 满足复杂环境和场景变化时目标跟踪的要求。

## 5 结论

本文提出了一种新的基于串行结构的多线索目标跟踪方法, 重在介绍基于串行结构的多线索目标跟踪框架, 并给出了一种具体的实现方法, 为使用多线索进行目标跟踪提供了一种新的思路。

实验结果表明, 该方法对目标跟踪有很高的鲁棒性的和准确性, 适用于对跟踪准确性可靠性都要求较高的场合。

基于串行结构的方法为使用多线索进行跟踪开辟了新的方向。与并行结构相比, 串行结构更符合人眼的跟踪机制, 其跟踪的准确性更高; 但串行结构中多线索自组织安排是难点, 多个线索如何有效、合理的使用有待进一步研究; 另外本文给出的跟踪方法的实时性不高, 因此如何减少计算量提高跟踪的实时性也是未来需要解决的问题之一。

## 参考文献 (References)

1 Comanicu D, Ramesh V, Meer P. Real-time tracking of non-rigid

- objects using mean shift [A]. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C], Hilton Head Island, South Carolina, USA, 2000:142 ~ 149.
- 2 Collins R T. Mean-shift blob tracking through scale space [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Madison, WI, USA, 2003: 234 ~ 240.
- 3 Lee Mun Wai, Cohen I, Soon Ki Jung. Particle filter with analytical inference for human body tracking [A]. In: Proceedings of IEEE Motion and Video Computing [C], Florida, USA, 2002:159 ~ 165.
- 4 Ozyildiz E, Krahnstover N, Sharma R. Adaptive texture and color segmentation for tracking moving objects [J]. Pattern Recognition, 2002, **35**(10): 2013 ~ 2029.
- 5 Baumberg A, Hogg D. An efficient method for contour tracking using active shape models [A]. In: Proceedings of IEEE Workshop on Motion of Non-rigid and Articulated Objects [C], Austin, Texas, USA, 1994:194 ~ 199.
- 6 Triesch J, Malsburg C. Self-organized integration of adaptive visual cues for face tracking [A]. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition [C], Grenoble, France, 2000:102 ~ 109.
- 7 Shen C, Hengel A, Dick A. Probabilistic multiple cue integration for particle filter based tracking [A]. In: Proceedings of the 7th International Conference on Digital Image Computing: Techniques and Applications [C], Sydney, Australia, 2003:399 ~ 408.
- 8 Itti L, Koch C. Computational modeling of visual attention [J]. Nature Reviews Neuroscience, 2001, **2**(3):194 ~ 230.
- 9 Wolfe J M. Guided search 2.0 a revised model of visual search [J]. Psychonomic Bulletin & Review, 1994, **1**(2): 202 ~ 238.
- 10 Friedman-Hill S R, Wolfe J M. Second-order parallel processing: Visual search for the odd item in a subset [J]. Journal of Experimental Psychology: Human Perception and Performance, 1995, **21**(3):531 ~ 551.
- 11 Zohary E, Hochstein S. How serial is serial processing in vision? [J] Perception, 1989, **18**(2): 191 ~ 200.
- 12 Neisser U. Cognitive Psychology [M]. New York, USA: Appleton Century Croft, 1967.
- 13 Gerhard Winkler. Image Analysis, Random Fields and Dynamic Monte Carlo Methods [M]. New York, USA: Springer, 1995:293 ~ 294.
- 14 Aheme F, Thacker N, Rockett P. The Bhattacharyya metric as an absolute similarity measure for frequency coded data [J]. Kybernetika, 1998, **34**(4): 363 ~ 368.
- 15 He D C, Wang L. Texture features based on texture spectrum [J]. Pattern Recognition, 1991, **24**(5):391 ~ 399.
- 16 Harvey Andrew C. Forecasting, Structural Time Series Models and the Kalman Filter [M]. Cambridge, UK: Cambridge University Press, 1989.